

Communication Theory and Physics

Transactions of the IRE Professional Group on Information Theory, v. 1 no. 1, 1953

by D. Gabor

Preface

This *IRE* paper is a second version of the paper with the same title which was first published as: “Communication Theory and Physics,” *Phil. Mag.*, ser. 7, vol. 41, pp. 1161–1187, 1950. This later version incorporated a number of typographical corrections, but the Appendices from the *Phil. Mag.* version were not included. Instead this paper provides references to the Appendices in the previous paper. In order to make this version self-contained, I have included the Appendices. A number of typographical errors, particularly in the equations, have been corrected.

It would appear from the publication dates that Gabor had about three years to think about and review the *Phil. Mag.* paper. But, this is not the case. The *Transactions of the IRE* of February 1953 was, in fact, devoted to the papers and discussion from the *Symposium on Information Theory*, London, September 1950. However, there is one substantive change in the *IRE* paper: the next to last paragraph before the References (pg. 1177) in the *Phil. Mag.* paper was removed. This paragraph first argues that “we can show at once that a light-amplifier for the reception of weak-light signals is impossible.” It seems that this was intended to refer to light *waves* and accurate measurement of amplitude and phase. The final sentence of the paragraph is: “This example demonstrates that *what we see* are always photons, not waves.” For completeness, I have included this paragraph here. It is appropriately shown as from the *Phil. Mag.* paper.

Subject to the point mentioned above, this version may be considered as a corrected version of the *Phil. Mag.* paper. Gabor usually referred to the *Phil. Mag.* paper in his subsequent publications.

In 1955 the *Transactions of the IRE Professional Group on Information Theory* was renamed the *IEEE Transactions on Information Theory*. Some issues of the *Transactions of the IRE* from 1953 to 1955 including v. 1, no. 1, are available on the *IEEE Xplore* web site. However, it is not possible to find these publications from the *IEEE Xplore* search. As one would expect, the right way is to Google on “Transactions of the IRE Professional Group on Information Theory.” This provides the address within the *IEEE Xplore* where the issues of the “Transactions of the IRE” are kept. Then select 1953 and the only volume and issue are vol. 1, i. 1. The paper is there. Unfortunately, this version was scanned using an 8.5 by 11 inch frame. The original publication is a reprographic copy of a typescript which was evidently prepared on Foolscap Folio (8 x 13in) paper, which was commonly used in England in the 1950’s. Therefore, material printed below about 11 inches is missing. (It is best to leave scanning to professionals.) This means that the paper is not conveniently available on the web in a fully readable form. This copy, reset in T_EX, and including the Appendices, is intended to provide web access to a complete copy of this fundamental work.

There is also a related report: “Lectures on Communication Theory,” *Research Laboratory of Electronics*, MIT Tech. Report No. 238, pp. 1-48, 1952. This written form of the Lectures, by Gabor, contains some of the material from these papers along with additional subjects. The lectures were given by Gabor at MIT during the Fall Term 1951. Thus, these lectures are after the Symposium which resulted in the the 1953 publication. This report refers to his *Phil. Mag.* paper and not the IRE version. It can be found at MIT through Google.

Finally, while I have proofread this draft, I would welcome any additional corrections that anyone may find. A [scanned version](#) of the original is also available on my web site.

Michael D. Godfrey
Statistics, Stanford University
email: michaeldgodfrey@gmail.com
Site: michaeldgodfrey/physics-information-communication

January 2012

Communication Theory and Physics

by D. Gabor
Imperial College, London

SUMMARY

The electromagnetic signals used in communication are subject to the general laws of radiation. One obtains a complete representation of a signal by dividing the time-frequency plane into cells of unit area and associating with every cell a “ladder” of distinguishable steps in signal intensity. The steps are determined by Einstein’s law of energy fluctuation, involving both waves and photons.

This representation, however, gives only one datum per cell, viz., the energy, while in the classical description one has two data: an amplitude and a phase. It is shown in the second part of the paper that both descriptions are practically equivalent in the long-wave region, or for strong signals, as they contain approximately the same number of independent, distinguishable data, but the classical description is always a little less complete than the quantum description. In the best possible experimental analysis the number of distinguishable steps in the measurement of amplitude and phase is only the fourth root of the number of photons. Thus it takes a hundred million photons per cell in order to define amplitude and phase to one percent each.

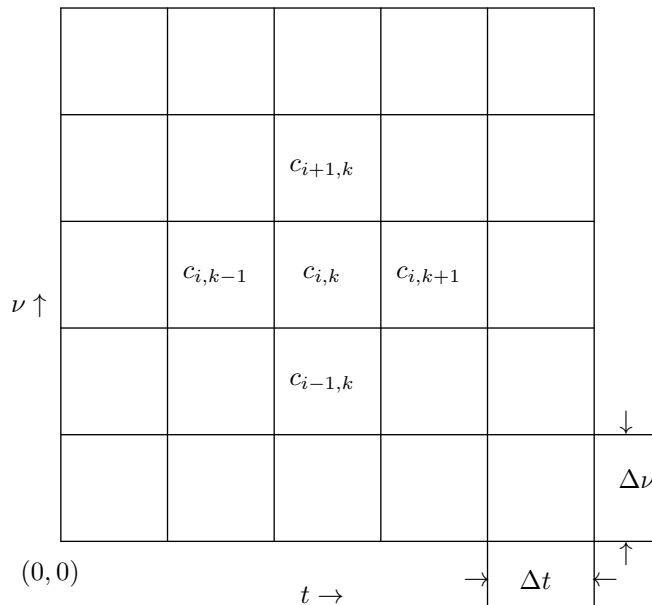
INTRODUCTION

COMMUNICATION theory has up to now developed mainly on mathematical lines, taking for granted the physical significance of the quantities which figure in this formalism. But communication is the transmission of physical effects from one system to another, hence communication theory should be considered as a branch of physics. Thus, it is necessary to embody in its foundations such fundamental physical data as the quantum of action, and the discreteness of electrical charges. This is not only of theoretical interest. With the progress of electrical communications toward higher and higher frequencies we are approaching a region in which quantum effects become all-important. Nor must one forget that vision, one of the most important paths of communication, is based essentially on quantum effects.

Some years ago, I have proposed a mathematical framework for the representation of signals (Gabor 1946 [1]). I have been rightly criticized for having left out noise, which is an essential feature of all communications. This will be remedied here, and at the same time the description will be brought in line with modern physics. But as the mathematical frame will serve as a useful foundation, it will be necessary to give first a short review of it.

(1) CLASSICAL REPRESENTATION OF SIGNALS

The previous work (Gabor [1, 2, 3]) started from the observation that the description of a signal in the conventional way, as a continuous function of time is redundant and non-physical. A continuous function contains in any interval, however small, an infinity of data, corresponding to an infinite range of frequencies. A similar objection can be raised against the Fourier representation, which involves infinite time. In the new description one considers the signal simultaneously as a function of frequency *and* of time. It is convenient to use only positive frequencies in the description. This can be done by introducing a certain complex function, whose real part is the physical signal. (The theory of these complex or “analytical” signals has in the meantime received interesting additions by the work of J. A. Ville (1948)[4]. If the time-frequency halfplane, Fig. 1, is divided, by any network, into cells of unit area, $\Delta t \Delta \nu = 1$, one finds that the signal in any domain containing a sufficiently large number of cells is fully described by associating two real data, or one complex datum with every cell. In other words, each cell has two degrees of freedom.

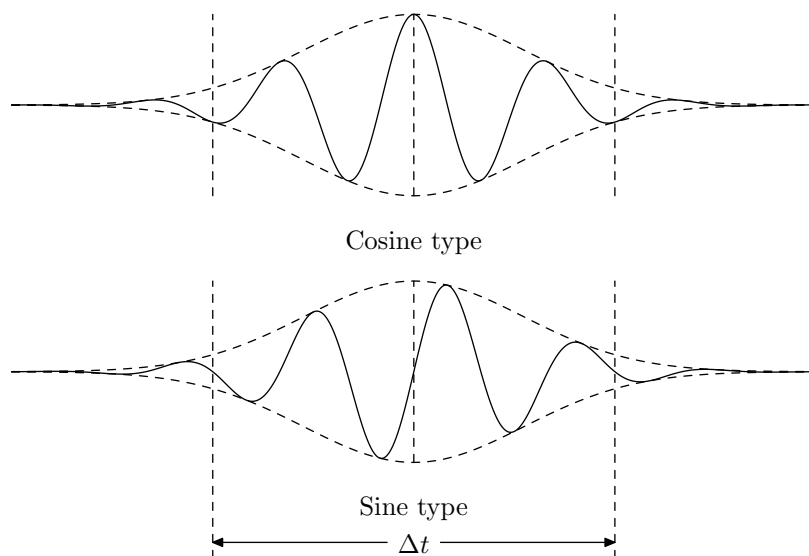
Fig. 1

Information diagram. The time-frequency half-plane is divided up into cells of unit area and an elementary signal is associated with each, with a coefficient c_{ik} .

One can represent an arbitrary signal in an infinity of ways as a linear function of certain “elementary” signals, associated with the individual cells. There is however one description of particular interest, in which the elementary signals are harmonic functions, modulated with a “Gaussian” signal. i.e. they have envelopes of probability shape (Fig. 2). These

share with other functions the property that their Fourier transforms have the same shape, but they are unique in the respect that the product of their “effective” duration and of their effective frequency width is the smallest possible for any function. Thus it can be said that these elementary functions overlap as little as possible. They have also the advantage that the familiar concepts “amplitude” and “phase” can be used in connection with them in the same way as with infinite harmonic functions. The complex elementary function is $\cos + j \sin$, if we denote by $\cos$ the even, and by $\sin$ the odd type of real elementary signal. Multiplying these with suitable complex coefficients c_{ik} , as indicated in Fig. 1, we can represent any arbitrary signal. Time-description and frequency-description (Fourier integral), can be considered as extreme special cases of this representation. In the first case the even elementary signal degenerates to a delta-function, and in the odd one to its derivative, in the second case both become ordinary harmonic functions.

Fig. 2



Elementary Signals of the “cosine type” (even) and of the “sine type” (odd).

The “matrix” representation, illustrated in Fig. 1, is proof against the objections raised against the pure “time” or “frequency” descriptions. but it does not go far enough. The infinity of data has been reduced to a finite number, but these data, i.e. the coefficients c_{ik} are still supposed to be *exactly* defined. But a single exact datum still contains an infinite amount of information, i.e. an infinite number of “yeses or noes.” In reality of course these amplitudes, like every physical datum, have a certain amount of uncertainty or “noise.” This has been taken into consideration in the mathematical theories of Shannon (1948)[5] and Tuller (1949)[6], where the noise amplitudes, or certain functions of them are assumed as known. But we cannot be satisfied with this in a physical theory. Even if all accidental imperfections of the instruments are eliminated, there remain certain basic uncertainties, which we are now going to investigate.

(2) STATISTICAL PROPERTIES OF THE INFORMATION CELL

In order to connect the mathematical scheme with physical reality we must first observe, that every physical signal has a certain energy associated with it. We can also associate a certain energy with every cell*, and for simplicity we will say that it is “contained” in the cell.

We will limit the discussion to electric communications, though most of our results will apply equally well to sound, or, in short, to communication by any quantity which is considered as *continuous* in classical physics. We observe first that all electric signals are conveyed by *radiation*. Even if lines or cables are used in the transmission, by the Maxwell-Poynting theory the energy can be located in empty space. Hence we can apply to our problem the well known results of the theory of radiation.

For simplicity we consider our communication system as having the uniform temperature T . The uncertainty connected with the concept of temperature will produce certain fluctuations in the energy of the cells, which we can calculate by the rules of statistical thermodynamics once we know the law for the mean thermal energy $\bar{\epsilon}_T$ of a cell, as a function of the temperature. This we obtain at once, if we observe that every cell in the signal has two degrees of freedom, of which only one counts for the purpose of statistics, as the other is of the nature of a phase. Thus, by Planck’s law

$$\bar{\epsilon}_T = \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1} \quad (1)$$

where ν is the mean frequency of the cell. In other words we identify every information cell with a “Planck oscillator.”

This requires perhaps a little more explanation, as physicists are less familiar with a discussion of radiation in terms of frequency and time than in terms of frequency and space. But the first case is immediately reduced to the second if we imagine the signal propagating with a velocity c , and plot the information diagram against the length ct instead of against the time t . The state of the field in such a linear system, (in which we consider one state of polarization only), can be represented by two systems of progressive waves in opposite directions, only one of which represents the signal in which we are interested. We count the degrees of freedom in this wave system, — which is what we have done in the last section, — and give it the energy $\bar{\epsilon}_T$ for each free amplitude, disregarding the phases. This is the application to the linear case of v. Laue’s well known derivation of Planck’s law for the radiation density by superposition of plane waves (von Laue 1914[7]).

* The energies of the elementary signals of the type discussed will not add up exactly to the energy of the whole signal, because they are not quite orthogonal, but the error will vanish in the mean over large numbers of cells. One can however, to meet objections, consider instead an orthogonal set of elementary signals, such as the signals with “limited spectrum,” introduced by Shannon (1949)[8] and by Oswald (1949, 1950)[9, 10] which have uniform spectral density inside and zero outside a frequency strip. This makes no difference to the following discussion, as no reference will be made to any special type of elementary signal.

The thermal energy $\bar{\epsilon}_T$ does not in itself represent “noise” i.e. an uncertain disturbing factor. It becomes disturbing only by its fluctuations. In order to obtain the mean square fluctuations of the energy, we apply Einstein’s formula

$$\overline{\delta\epsilon_T^2} = \overline{(\epsilon_T - \bar{\epsilon}_T)^2} = kT^2 \frac{d\bar{\epsilon}_T}{dT} \quad (2)$$

which gives

$$\overline{\delta\epsilon_T^2} = h\nu\bar{\epsilon}_T + \bar{\epsilon}_T^2 \quad (3)$$

Einstein’s interpretation of this equation is well known (*cf.* Born 1949[11], p. 81). The second term has been identified by H. A. Lorentz with the fluctuations due to the interference of waves with random phases*. But, the first term suggests that the energy is concentrated in light quanta or photons of energy $h\nu$ which fluctuate in any element as if they were independent particles. As both effects are present simultaneously, and are acting as if they were independent of one another, it is not possible to make a simple physical picture of the process. But fortunately this is not necessary. Applying Einstein’s argument to the case when a signal is present, so that the mean energy $\bar{\epsilon}$ in the cell exceeds the thermal mean energy $\bar{\epsilon}_T$ we obtain in Appendix I. (from Gabor[12])

$$\overline{\delta\epsilon^2} = \overline{(\epsilon - \bar{\epsilon})^2} = h\nu\bar{\epsilon} + 2\bar{\epsilon}\bar{\epsilon}_T - \bar{\epsilon}_T^2 \quad (4)$$

Expressing the energy by the number of photons N in the cell, so that $\epsilon = Nh\nu$, $\bar{\epsilon}_T = \bar{N}_T h\nu$, this becomes

$$\overline{\delta N^2} = \overline{(N - \bar{N})^2} = \bar{N}(1 + 2\bar{N}_T) - \bar{N}_T^2 \quad (5)$$

with

$$\bar{N}_T = \frac{1}{e^{\frac{h\nu}{kT}} - 1}. \quad (6)$$

$\bar{N}_T$, the number of “thermal” photons per cell, is a large number for the frequencies used in electrical communications. One can use the approximate formula

$$\bar{N}_T = \frac{kT}{h\nu} = \frac{kT}{hc} \lambda = 0.7\lambda T$$

* In technical theories of thermal noise it is usually forgotten that it is not the noise power but its fluctuations which cause the disturbance. But if the quantum effect is small and the second term in eq.(3) predominates, the r.m.s. value of the noise power fluctuation is equal to the noise power itself, hence this error is without consequences.

From eq.(3), neglecting the quantum term one can easily derive Nyquist’s well known rule (Nyquist 1928)[13] that the resistance R can be considered as containing a “noise generator” with mean square electromotive force

$$\overline{E^2} = 4kTR\Delta\nu.$$

The proof can be given in the same form as Nyquist has done, by substituting a cable with wave impedance R for the resistance. But one must add the condition that not only the noise power, but also its fluctuations follow the same rule in the resistance as in the cable. This is no arbitrary rule, the necessity of a uniform law of mean square fluctuations follows directly from the second principle, as Szilard (1925)[14] has shown.

which gives $\bar{N}_T = 210$ for a wavelength of 1cm and a temperature of 300°K . On the other hand for visible light, $\lambda = 5.10^{-5}\text{cm}$, the approximate formula

$$\bar{N}_T = e^{-\frac{h\nu}{kT}} = e^{-\frac{1}{0.7\lambda T}}$$

gives $\bar{N}_T = e^{-94} = 10^{-39}$. Hence for visible light, at ordinary temperatures there is practically no thermal noise, and the fluctuation becomes pure “quantum noise,” which follows the law $\delta N^2 = \bar{N}$.

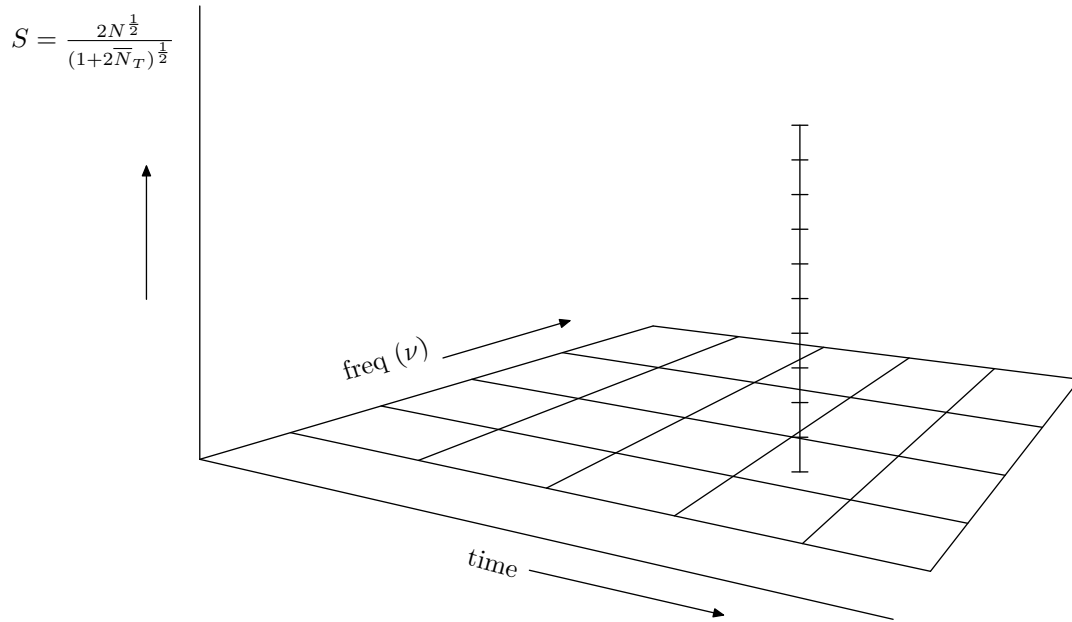
These results enable us to construct a complete physical representation of a signal. We see that the state of an information cell is completely determined by the stochastic number N , the number of photons. We can now construct a scale or “ladder” of distinguishable states, on which every step corresponds to a reasonably ascertainable difference. It is an evident suggestion to adopt the r.m.s. fluctuation of N as the unit step.* With this convention, the number of steps distinguishable below a maximum level N_m is, by eq.(5)

$$\begin{aligned} S &= \int_{\bar{N}_T}^{N_m} \frac{dN}{(\delta N^2)^{\frac{1}{2}}} = \int_{\bar{N}_T}^{N_m} \frac{dN}{\left[N(1 + 2\bar{N}_T) - \bar{N}_T^2\right]^{\frac{1}{2}}} = \\ &= \frac{2}{1 + 2\bar{N}_T} \left(\left[N_m(1 + 2\bar{N}_T) - \bar{N}_T^2 \right]^{\frac{1}{2}} - \left[\bar{N}_T(1 + \bar{N}_T) \right]^{\frac{1}{2}} \right) \approx \frac{2N_m^{\frac{1}{2}}}{(1 + 2\bar{N}_T^{\frac{1}{2}})} \end{aligned} \quad (7)$$

The last formula is valid for large signals.

* By the theorem of Bienaymé and Tchebycheff the probability of an error k times exceeding the r.m.s. error or fluctuation is smaller than $1/k^2$, whatever the law of the fluctuations may be.

Fig. 3



Representation of an arbitrary signal. A ladder or “proper scale” of distinguishable values is erected in every cell, on which the occupation is marked off.

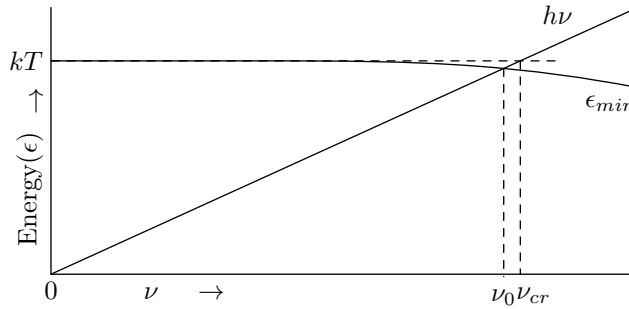
In the useful terminology introduced by D. M. MacKay (1950)[15], S is the “proper scale” of the photons. Fig. 3 illustrates the representation of a signal in three dimensions, with the photon scale at right angles to the time-frequency plane. If instead we plotted $\log_2 S$, the ordinates would give directly the equivalent number of binary selections or “bits.” This is the number of “yeses or noes” required to fix the position of the signal on the ladder.

It is of interest to inquire about the minimum energy required for the transmission of the first “bit” of information. By our convention this cannot be less than the r.m.s. value of the thermal energy fluctuations, though it can be more. Combining eqs. (1) and (3) one obtains

$$\epsilon_{min} = (\overline{\delta\epsilon_T^2})^{\frac{1}{2}} = \frac{h\nu}{2 \sinh \frac{h\nu}{2kT}}. \quad (8)$$

But on the other hand this energy cannot be less than one quantum, $h\nu$. Thus the energy required for the first “bit” is either ϵ_{min} as given by eq. (8), or $h\nu$, whichever is the larger of the two. As shown in Fig. 4, the two lines cross at $h\nu_0 = 0.96kT$, which is only slightly less than kT . Thus up to a critical frequency ν_{cr} an energy kT is sufficient for the first step, but no communication is possible with an energy of less than kT . The interesting feature of this result is its generality, it applies to the unknown processes in the nervous system as well as to electrical communications.

Fig. 4



The energy required for the transmission of the first “bit” of information.

(3) STATISTICS OF SIGNALS

Up to this point we have directed our attention only to one cell. A short discussion of signals and of ensembles of signals may not be out of place before we return to the physical analysis of our results.

A signal is a system of, say, N cells, preferably but not necessarily contiguous in the information plane. We consider a large number of such systems in a stationary transmission, in which they differ only by their position in time, not in frequency, and we speak of this number as of an “ensemble.” Evidently the analogy with statistical mechanics is not very perfect. In statistical mechanics we can either follow a system of an ensemble over a long time, or we can look at all systems in the ensemble simultaneously, while here we can take only the first view. It is also somewhat questionable whether the often used expression “ergodic” is justified. In its original sense, due to Willard Gibbs, it means that each system (each signal) spends equal times in all states compatible with a given energy, which is not true for most stationary transmissions usually considered as “ergodic” in communication theory. Thus we prefer to avoid this term.

The most important mean value in such an ensemble is that of the *entropy*. In order to clarify the connections between physics and communication theory, it may be useful to consider this problem in two stages. In the first we consider all configurations of the system as equally probable, and define the entropy $k \log P$, where P is the number of all these possible configurations. In the second stage, however, we give them different probabilities or “weights.” The first stage is in close connection with physics, actually it is the problem of calculating the entropy of an “ergodic” system, in the original meaning of the word. According to quantum statistics all simple, accessible states have equal probability, and all the levels of Planck oscillators are of course simple. (cf. Jordan 1933[16])

As an example let us estimate the mean entropy of a system of N cells in a transmission in which the mean energy level is $\bar{S}^2$, and the r.m.s. deviation from the mean is ΔS_n^2 , for brevity. (The suffix n had to be added, as the standard deviation is dependent on the size of the sample.) For simplicity we assume $\Delta S_n^2 \ll \bar{S}^2$, i.e. a small degree of “modulation.” This

allows us to equate the number P of possible states to the number of points with positive, integer coordinates inside a spherical shell in n -dimensions, whose mean radius R is given by

$$R^2 = n\overline{S^2},$$

while its thickness is

$$2\Delta R = n\Delta S_n^2/R.$$

Well known formulas for the volume of an n -dimensional sphere give

$$\begin{aligned} P &= \frac{(\pi/4)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n+1)} nR^{n-1} 2\Delta R \\ &= \frac{(\pi/4)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n+1)} n(n\overline{S^2})^{\frac{1}{2}n} \frac{\Delta S_n^2}{\overline{S^2}} \approx \frac{1}{\sqrt{\pi}} \left(\frac{1}{2}e\pi\overline{S^2}\right)^{\frac{1}{2}n} \frac{\sqrt{n}\Delta S_n^2}{\overline{S^2}} \end{aligned} \quad (9)$$

We have used Stirling's formula to approximate the Γ function for large arguments. For sufficiently large n , however complicated the signals, they must behave as if they were independent, and we must have asymptotically

$$n(\Delta S_n^2)^2 \rightarrow \text{const.}$$

hence, apart from a constant,

$$S \rightarrow kn \log \left(\frac{1}{2}e\pi\overline{S^2} \right)^{\frac{1}{2}}, \quad (10)$$

a formula also obtained by Shannon (1949)[8], though our notations are somewhat different.

Apart from the factor k the entropy (10) is also the measure of the quantity of information, in accordance with its definition as the logarithm of the number of possible, equally probable selections. The connection between entropy in the thermodynamic meaning of the word, has been cleared up by Szilard (1929)[17] who proved the remarkable theorem that information corresponding to an s -fold selection enables the receiving system to reduce the entropy of the transmitter by a *maximum* of $k \log s$. He proved also that any mechanism acquiring this information must increase the entropy by a minimum of $k \log s$, in accordance with the second principle.

So far the concepts of information and of entropy are closely parallel, even identical. But a somewhat new feature was introduced by Shannon (1948)[5], who defined the *mean entropy per symbol* in a transmission in which the symbols have probabilities, i.e. relative frequencies p_i as

$$H = -k \sum p_i \log p_i \quad H > 0. \quad (11)$$

This new concept is at an appreciable remove from the physical entropy previously discussed. The probabilities p_i have no direct relation to the structure of the signal, they are determined by the source. A symbol can be represented by any configuration, or group of configurations in a basic group of cells, provided that the basic group allows at least as many distinguishable configurations as there are symbols, and even this fairly general definition does not exhaust the almost unlimited possibilities of coding. Hence the interesting properties of the expression (11) demonstrated by Shannon, must be attributed to its mathematical form rather than to its intrinsic relation with the physical concept going by the same name.

(4) THE CONNECTION BETWEEN QUANTUM REPRESENTATION AND CLASSICAL DESCRIPTION

Comparing the representation of a signal, illustrated in Fig. 3, with the classical, mathematical description, it appears that we have lost something. Previously we had two data per cell, we are left now with one only, a function of the quantum number N , which corresponds to the energy. What has happened to the phase?

It is not difficult to give an answer to this question on general lines. N is not an exact datum but a stochastic number, which represents a finite amount of information. But between one datum of finite accuracy and two there is not transcendental abyss of the kind which exists between a simple and a twofold infinity. One finite datum can very well contain two independent, finite data, provided that their aggregate information is not more than the original. It will be shown that this is indeed the case, and that the maximum amount of information on amplitude and phase are both contained in the single photon scale.

An electromagnetic signal can be physically analyzed in various ways, but it will be instructive to consider first extreme cases only. One extreme is a *counter*, an instrument which records single photons. It follows immediately from Heisenberg's uncertainty principle that the "time resolution" of such an instrument cannot be better than a whole cycle, hence the phase remains entirely unobservable. The other extreme is any classical field-measuring instrument, capable of recording the electromagnetic field in the signal as a function of time. It may be called for brevity a proportional *amplifier*, as amplification of weak signals is one of its essential functions. There is no need to consider intermediate instruments, as it will be seen that in a certain range of very weak signals every classical amplifier operates as a sort of "proportional counter."

One limit for the operation of any such instrument is given by the well known uncertainty relation of the quantum theory of radiation, (*cf.* Heitler (1944)[18] p. 68).

$$\Delta N \Delta \phi \geq 1 \quad (12)$$

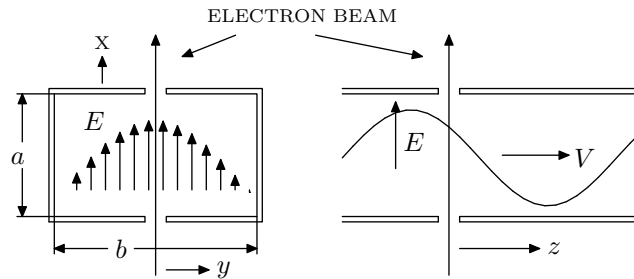
where ΔN is the uncertainty in the photon number N , and $\Delta \phi$ the uncertainty in the phase ϕ . But this gives merely an upper limit in our case, as we want to determine amplitude and phase *simultaneously*, each as accurately as possible. It will be seen that in fact the limiting accuracy is much below what might be expected from the inequality (12).

We will approach the problem by a detailed analysis of a particular type of proportional detector-amplifier. Subsequently we will try to improve its performance to the extreme limit. It will then be found that all special features of the device vanish from the formulas, thus the final result can be considered as of general validity.

Assume that the electromagnetic signal is introduced into a rectangular wave guide (Fig. 5) in the TE₀₁ mode, i.e. with an electric field in the x -direction, independent of the x -coordinate, which is at right angles to the direction of propagation z . An electron beam passes through two small holes along the x -axis, in the plane of maximum intensity, with velocity v and current intensity J . The alternating acceleration and retardation of the

electrons produces velocity modulation in the beam, which by well known methods can be translated into current modulation, i.e. producing an alternating component superimposed on the mean value J . One method utilizes the “bunching” of the electrons which takes place at a certain distance from the guide, but one can also deflect the beam by a constant field and make it play between two collecting electrodes, close together. In either case one obtains an alternating current which in first approximation is proportional to the d.c. current and to the relative accelerations and retardations suffered by the electrons in the wave guide.

Fig. 5



Analysis of an electromagnetic signal in a wave guide by an electron beam.

Let us measure the energy exchange between the electrons and the field in quantum units, $h\nu$, where ν is the mean frequency of the signal. (It will soon be seen that in order to make the exchange intense, the waveband of the signal must be made so narrow that it is permissible to take the arithmetic mean.) Let $\bar{N}$ be the mean number of photons in an information cell, which passes through a cross-section of a guide in a time $\Delta t = 1/\Delta\nu$. During this time $\bar{N} = J\Delta t/e$ electrons pass through it, and exchange in the mean $\bar{n}$ quanta with the field, either by losing or gaining energy. The positive number $\bar{n}$ is the essential parameter of the process. If $\bar{n}$ is a large number, which is possible only if $\bar{N}$ is also large, the interchange will be essentially classic, if $\bar{n}$ is small quantum phenomena will dominate.*

The detailed calculations may be found in the Appendices II. and III. (from Gabor (1950)[12]), only the results will be discussed here. The first contains the classical theory of exchange, valid for large $\bar{N}$ and $\bar{N}$, the second a wave-mechanical calculation valid for small exchange parameters. In the classical theory the result is

$$\bar{n}^2 = \frac{32}{\pi^3} \left(\frac{2\pi e^2}{hc} \right) \frac{\sin^2 \theta}{\theta} \frac{v}{\nu b} \left[1 - \left(\frac{c}{2b\nu} \right)^2 \right]^{-\frac{1}{2}} \frac{\Delta\nu}{\nu} \bar{N} \quad (13)$$

Thus the exchange parameter is proportional to the square root of the photon number, as may be expected. Of the dimensions of the waveguide, the width b appears explicitly in the

* Quantal energy exchange between electrons and the field in a wave guide at high quantum numbers has been previously discussed by Lloyd P. Smith (1946)[19], but we cannot agree with most of his results. Monokinetic electrons and exchange of sharply defined quanta on the one hand, well defined entrance phases and short transit times on the other are mutually exclusive phenomena by the Uncertainty Principle, hence we believe that only certain averages over Smith's detailed results have physical significance.

factor of $\bar{N}$, while the depth a is contained in the angle

$$\theta = \pi \nu a / v \quad (14)$$

which is one-half of the “transit angle.” c is the velocity of light. In addition there appears a factor, which is the reciprocal of

$$\frac{hc}{2\pi e^2} = 137$$

the fundamental number which connects photons with electrons.

Anticipating that the best measurement will require an intense interchange, we now try to increase the coefficient of $\bar{N}$ in eq.(13) by all available means. First we make the factor $\sin^2 \theta / \theta$ a maximum. This is 0.723 and is obtained with $\theta = 67^\circ$, i.e. at a transit angle of 134° . This disposes of the depth a of the waveguide. The optimum width b is determined by the condition that the group velocity

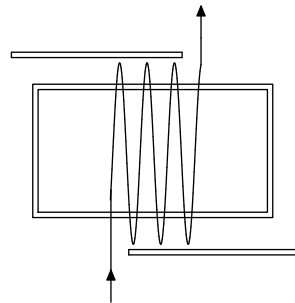
$$U = c \left[1 - \left(\frac{c}{2b\nu} \right)^2 \right]^{\frac{1}{2}}$$

must be as small as possible. But the smallest value is reached when U vanishes at the low-frequency limit, $\nu - \frac{1}{2}\Delta\nu$, of the band. Substituting these values into eq.(14) one obtains in the optimum case

$$\bar{n}^2 = \frac{1.5}{137} \left(\frac{v}{c} \right) \left(\frac{\Delta\nu}{\nu} \right)^{\frac{1}{2}} \bar{N} \quad (15)$$

All special features of the device have vanished in this formula, apart, perhaps, from the unimportant factor 1.5. But it is evident that the factor of $\bar{N}$ must be always much smaller than unity, while its best value, as will be shown later, is just unity. There exists, however, a further possibility for improving its performance. Assume that we can make each electron perform repeated passages through the guide, each transit, in the opposite direction exactly half a cycle after the last. (See Fig. 6) (This is possible in principle, as the optimum transit angle is about 134° .)

Fig. 6



Repeated Passages

If the frequency were known beforehand, the number of passages would be limited only by the consideration that by repeated gains or losses the electron would be bound to fall out of

synchronism. But if the signal had a single frequency, known in advance, there would be of course no communication. However, even if the frequency is known beforehand only within $\nu \pm \frac{1}{2}\Delta\nu$, one can make the number P of passages as great as $\nu/\Delta\nu$, without risking a phase error of more than $\pm\frac{1}{2}\pi$, and it can be shown that these passages are almost of equal value, so that $\bar{n}$ is increased very nearly by a factor P . The number of passages required to make the coefficient of $\bar{N}$ in eq.(15) unity is

$$P_{opt} = 9.5 \left(\frac{c}{v}\right)^{\frac{1}{2}} \left(\frac{\nu}{\Delta\nu}\right)^{\frac{1}{4}} \quad (16)$$

i.e. at least of the order of ten. But this number must not exceed $\nu/\Delta\nu$, hence we obtain the condition that in order to realize optimum conditions the frequency band must be so restricted that

$$\left(\frac{\nu}{\Delta\nu}\right) > 20.2 \left(\frac{c}{v}\right)^{\frac{2}{3}} \quad (17)$$

This is a somewhat surprising result. In the mathematical representation it did not matter whether we divided up the frequency band into broad or narrow strips. But by the intervention of the number 137, ($20.2 = (137/1.5)^{2/3}$), it turns out that only narrow frequency bands are capable of accurate analysis by means of electrons. (If ions with charge Z were used one would have to replace 137 by $137/Z^2$, and the condition would be less stringent.)

Thus, at least in theory, the device could be perfected up to the optimum performance. The practical difficulties are of course evident. In practice one would rather replace the wave guide by a “high-Q” resonator, but this would somewhat complicate the theory.

We thus find, assuming P passages, in the “classical” case

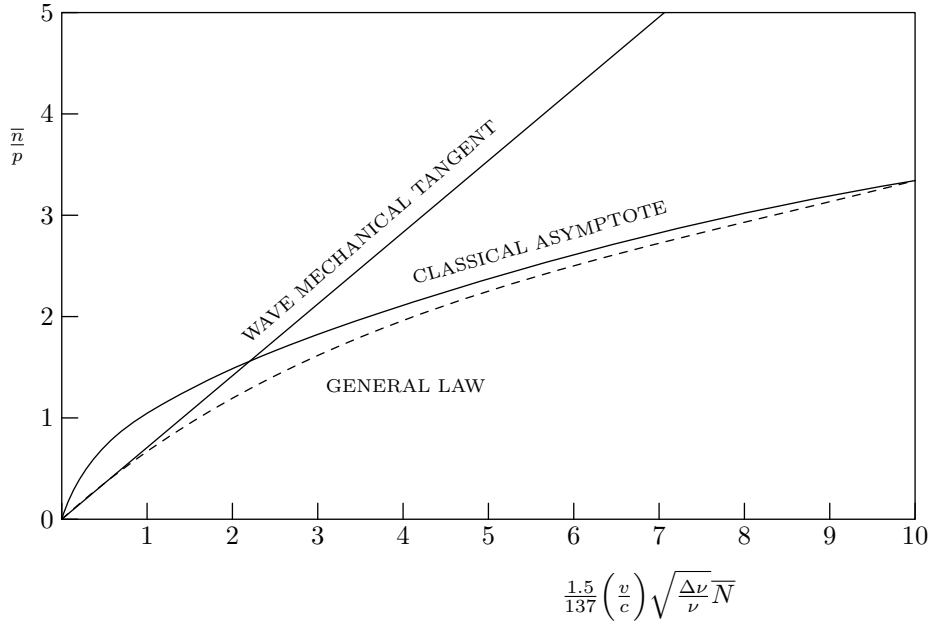
$$\left(\frac{\bar{n}}{P}\right)^2 = \frac{1.5}{137} \left(\frac{v}{c}\right) \left(\frac{\Delta\nu}{\nu}\right)^{\frac{1}{2}} \bar{N} \quad (18)$$

while the corresponding wave-mechanical formula, valid for very weak interchange is,

$$\frac{\bar{n}}{P} = \frac{1.0}{137} \left(\frac{v}{c}\right) \left(\frac{\Delta\nu}{\nu}\right)^{\frac{1}{2}} \bar{N} \quad (19)$$

Apart from the factor $2/3$ the coefficient of $\bar{N}$ is the same in both cases, but this time the quantum exchange $\bar{n}$ is proportional to the photon number itself, not to its square root. It can be said therefore that *for small photon concentrations the device acts as a counter, at large concentrations as a field measuring instrument.* The intermediate region is difficult to calculate, but as shown in Fig. 7, the two branches can be connected by a plausible curve.

Fig. 7



Mean Quantum Exchange between Electrons and Photons. The initial tangent is calculated from wave mechanics, the asymptotic parabola classically.

One can also interpret the results in this way: If there are few photons present, there will be few collisions, and equal probabilities of gains and losses, at any instant. With increasing photon concentration repeated collisions will increase in number, and the resulting loss or gain increases with the square root of the photon number only; but this resultant has now a prevailing direction, which changes its sign with the frequency of the signal. At this stage the “classical field” has developed.

Having ascertained that, with certain reservations, we can make the exchange as strong as we like, we ask the question: If we know the average photon number $\bar{N}$, how must we adjust the electron beam current J , and the exchange parameter $\bar{n}$ in order to measure the field amplitude E with maximum accuracy? And having made these adjustments, how many steps shall we be able to distinguish in the scale of the field amplitudes? Evidently this question has a precise meaning only in the “classical” range of large $\bar{n}$ and $\bar{N}$, and the following considerations relate only to this case. In order to simplify the problem we neglect the thermal noise, i.e. we put $\bar{N}_T = 0$, so that the relative accuracy on the photon scale would be $1/\bar{N}^{\frac{1}{2}}$, and the total number of steps in the photon ladder $2\bar{N}^{\frac{1}{2}}$. The calculations are carried out in Appendix IV (from Gabor 1950[12]), here we give only the physical considerations.

The quantity to be measured is the electric amplitude in the information cell, which is proportional to the square root of the photon number. The measured quantity, on the other hand, is the alternating electron current, which, as mentioned above, is proportional to $\bar{n}\bar{J}$ or $\bar{n}\bar{M}$ for not too strong signals, $\bar{M}$ being the mean number of exploring electrons per cell.

For the optimum we impose the condition that the relative mean square deviation of the quantity measured from the quantity to be measured shall be as small as possible, i.e.

$$\frac{\overline{(nM - CN^{\frac{1}{2}})^2}}{(\overline{nM})^2} = \min. \quad (20)$$

where the proportionality factor C is determined from the condition

$$\overline{(nM - CN^{\frac{1}{2}})} = 0$$

We have to choose $\bar{n}$ and $\bar{M}$ so as to satisfy the condition (20) for given $\bar{N}$. There must be an optimum, for these reasons: A too weak interchange $\bar{n}$ would leave the cell unexplored. A too strong interchange on the other hand will *interfere with the object of the measurement* and spoil it. Though in the mean electrons are as often accelerated as retarded, fluctuations in the numbers M_1 and M_2 of electrons which pass through accelerating and retarding phases might produce extra photons, which could not be distinguished from those belonging to the signal, or annihilate some. The spurious photons are generated according to a law

$$\overline{\delta_2 N} = \bar{n}(M_1 - M_2) = \bar{n}(\delta M_1 - \delta M_2) \quad (21)$$

as $M_1 = M_2 = \frac{1}{2}\bar{M}$. We have written $\overline{\delta_2 N}$ for this number, considering it as a fluctuation which must be added to the natural fluctuation $\delta_1 N$, whose law is $\overline{\delta_1 N^2} = \bar{N}$. The two fluctuations must be considered as independent.

It is already evident from the above that the fluctuations in the number of beam electrons i.e. the “shot effect” plays an important part in these phenomena. A too weak current has a high relative fluctuation. A too strong current, especially aided by a large exchange will again spoil the object. It may be noted that *we have here a type of uncertainty which springs directly from the fact that photons and electrons are discrete*, without any reference to the physical values of h and of e .

The relative error according to eq.(20) is calculated on the basis of eq.(21), together with such evident assumptions as the independence of the fluctuations of n and M and the “natural” part of δN . We assume also “normal” shot effect, $\overline{\delta M^2} = \bar{M}$. The result is

$$\frac{\overline{(nM - CN^{\frac{1}{2}})^2}}{(\overline{nM})^2} = \frac{1}{4\bar{N}} + \frac{1}{\bar{n}} + \bar{n}\left(\frac{1}{\overline{nM}} + \frac{\overline{nM}}{4\bar{N}^2}\right) \quad (22)$$

This is a minimum for

$$\overline{nM} = \bar{n}\bar{M} = 2\bar{N} \quad \overline{n^2} = \bar{N} \quad (23)$$

from which $\bar{M} = 2\bar{N}^{\frac{1}{2}}$. This gives the simple rule that, for optimum analysis of the signal, one must take one electron for every step in the scale of photons, and the interchange $\bar{n}$ must be itself equal to one distinguishable step at the level $\bar{N}$. This again is a general rule, quite independent of the special model from which we started.

Substituting these values, we obtain for the minimum of the mean square relative error in the measurement of amplitudes

$$\left(\frac{(nM - CN^{\frac{1}{2}})^2}{(nM)^2} \right)_{min} = \frac{1}{4N} + \frac{2}{n} = \frac{1}{4N} + \frac{2}{N^{\frac{1}{2}}} \quad (24)$$

As we are dealing with large photon numbers only, the first term at the right hand side can be neglected with respect to the second. Thus we see, applying the same rule which we have used in constructing the photon ladder, that the smallest distinguishable relative step in the amplitude scale is $\sqrt{2}$ times larger than the square root of the corresponding quantity in the photon scale. In other words, *the proper scale of the amplitudes will contain always less than the square root of the number of steps in the photon scale*, apart from the factor $\sqrt{2}$ for the reason that the optimum setting is of course possible for one level only.

Having determined the proper scale of the amplitudes, a simple application of the Uncertainty Principle, shows that the proper scale of the phase must also contain less steps than the square root of the photon scale. Thus, summing up, we see that the classical description of the signal, by being too detailed, gives in fact a somewhat *smaller total amount of information* than the quantum description.

The classical method of description, though theoretically inferior, may of course be still the best practically in the range of frequencies use for electrical communications, where efficient photon counters are not available. Conditions are different in theoretical region, where detectors of the counter type – such as the eye – are not far from perfection. In this region analysis in terms of electromagnetic waves is as yet technically impossible, but it is interesting to note that even if it were possible, it would not be very practical. Progress in the field of microwaves is now actually approaching a region where the two different methods of analysis may become competitive. We see from our results that it takes about a hundred million photons per information cell in order to define amplitude and phase of the signal to 1% each. Remembering that at 1cm wavelength the number of thermal photons per cell is only about 200, it may be seen that the time may be not far off when the imperfections of the classical method of description will manifest themselves even in electrical communications.

The following paragraph did not appear in this version. It only appeared in the *Phil. Mag.* version. It is included here for completeness.

Will it ever be possible to generate and to amplify light by an extension of microwave techniques? Generation may not be unthinkable, but we can show at once that a light-amplifier for the reception of weak-light signals is impossible. It follows from our results that an electromagnetic signal in which amplitude and phase are defined with a relative accuracy ϵ must carry at least a power

$$h\nu^2\epsilon^{-4}(\Delta\nu/\nu). \quad (25)$$

Applied to visible light with $\lambda = 5000\text{\AA}$, a line width of 1Å. and $\epsilon = 0.01$, we obtain 4.7 watts. Even if amplitude and phase were defined to 10 per cent only,

the power would still be 0.0005 watts, a very high intensity in terms of visible light. This example demonstrates that *what we see* are always photons, not waves.

It may be hoped that these considerations have shown that the concepts of information theory may well prove their usefulness when applied to problems of physics.

References

- [1] Gabor, D.: Theory of Communication. *J IEE* 93, pt. III, 429–457 (1946)
- [2] Gabor, D.: New Possibilities in Speech Transmission. *J IEE* 94, pt. III, 369–390 (1947)
- [3] Gabor, D.: Acoustical Quanta and the Theory of Hearing. *Nature* v. 159, 591–594 (1947)
- [4] Ville, J.: Théorie et Applications de la Notion de Signal Analytique. *Câbles et Transmissions* 2, no. 1, 61–74 (1948)
- [5] Shannon, C. E.: A Mathematical Theory of Communication. *Bell Syst Tech J* 27, nos. 3, 4, 379–423, 623–656 (1948)
- [6] Tuller, W. G.: Theoretical Limitations on the Rate of Transmission of Information. *Proc IRE* 37, no. 5, 468–78 (1949)
- [7] M., L.: Die Freiheitsgrade von Strahlenbündeln. *Ann d Phys* 44, 16, 1197–1212 (1914)
- [8] Shannon, C. E.: Communication in the Presence of Noise. *IRE* 37, no. 1, 10–21 (Jan., 1949. Reprinted in *Proc. IEEE*, 86, no. 2, pp.447–457, 1998)
- [9] Oswald, J.: Sur Quelques Propriétés des Signaux à Spectre Limité. *J Comptes Rendus* 229, 21–22 (1949)
- [10] Oswald, J.: Les Signaux à Spectre Limité et leurs Transformations. *Cables et Transmissions* 4, i. 3, 197–215 (1950)
- [11] Born, M.: *Natural Philosophy of Cause and Chance*. Dover, New York (1949). Dover edn. 1964
- [12] Gabor, D.: Communication Theory and Physics. *Phil Mag* 41, ser. 7, no. 322, 1161–1187 (1950)
- [13] Nyquist, H.: Thermal Agitation of Electric Charge in Conductors. *Phys Rev* 32, 110–113 (July, 1928)
- [14] Szilard, L.: Über die Ausdehnung der phänomenologischen Thermodynamik auf die Schwankungerscheinungen. *Z Phys* 32, 753–788 (1925)
- [15] Mackay, D. M.: Quantal Aspects of Scientific Information. *Phil Mag* s. 7, v. 41, 289–311 (1950)
- [16] Jordan, P.: *Statistische Mechanik auf Quantentheoretischer Grundlage*. Fr. Vieweg, Braunschweig (1933)
- [17] Szilard, L.: Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Z Phys* 53, no. 11-12, 840–856 (1929)
- [18] Heitler, W.: *The Quantum Theory of Radiation*. Oxford, UK (2nd. edn. 1944)

- [19] Smith, L. P.: Quantum Effects in the Interaction of Electrons with High Frequency Fields and the Transition to Classical Theory. *Phys Rev* 69, no. 5/6, 195–210 (1946)
- [20] Carson, J. R. and Fry, T. C.: Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency-Modulation. *Bell Syst Tech J* 16, i4, 513–540 (1937)
- [21] Macdonald, D. K. C. and Kompfner, R.: Fluctuation Phenomena Arising in the Quantum Interaction of Electrons with High-frequency Fields. *Proc Inst Rad Engrs* 37, No. 12, 1424–1426 (1949)

APPENDIX I.

Energy Fluctuations in an Information Cell in the Presence of a Signal.

As in the case of thermal equilibrium, the mean square energy fluctuation is the sum of a classical and a quantum term.

The classical term./ Consider, for simplicity, a (complex) Fourier component E_S of the signal, and a component E_T of the thermal noise, corresponding to two different (circular) frequencies, say, ω_1 and ω_2 . The instantaneous energy density, resulting from the interference, is proportional to

$$E_S E_S^* + E_T E_T^* + \left(E_S E_T^* \exp[j(\omega_1 - \omega_2)t] + E_S^* E_T \exp[j(\omega_1 - \omega_2)t] \right),$$

where the asterisks stand for conjugate complex values. The first two terms are the energy of the signal and the energy of the noise, the rest arises from interference. (Beats.) Let us write

$$\epsilon = \epsilon_S + \epsilon_T + \epsilon_{ST}.$$

The mean value of the interference energy ϵ_{ST} , is *nil*, but its mean square is

$$\overline{\epsilon_{ST}^2} = 2 \overline{E_S E_S^*} \overline{E_T E_T^*} = 2 \bar{\epsilon}_S \bar{\epsilon}_T.$$

Using this and the relations

$$\overline{\epsilon_S \epsilon_{ST}} = \overline{\epsilon_T \epsilon_{ST}} = 0$$

which are evident, as there is no correlation between the signal and the noise, one obtains

$$\overline{(\epsilon - \bar{\epsilon})^2} = \overline{[(\epsilon_S + \epsilon_T + \epsilon_{ST}) - (\bar{\epsilon}_S + \bar{\epsilon}_T)]^2} = \overline{\epsilon_T^2} - \bar{\epsilon}_T^2 + 2\epsilon_S \bar{\epsilon}_T.$$

One knows the first two terms from Lorentz's calculation, which gives, in the absence of a signal

$$\overline{\epsilon_T^2} - \bar{\epsilon}_T^2 = \bar{\epsilon}_T^2,$$

hence

$$\overline{(\epsilon - \bar{\epsilon})^2} = 2\epsilon_S \bar{\epsilon}_T + \bar{\epsilon}_T^2,$$

or, as

$$\begin{aligned} \epsilon_S &= \bar{\epsilon} - \bar{\epsilon}_T, \\ \overline{(\epsilon - \bar{\epsilon})_{\text{cl.}}^2} &= (2\bar{\epsilon} - \bar{\epsilon}_T) \bar{\epsilon}_T \end{aligned} \tag{I.1}$$

This is the classical part of the fluctuation, arising from the interferences of waves. *The quantum term* is always given by the “law of rare events”

$$\overline{(\epsilon - \bar{\epsilon})_{\text{qu.}}^2} = h\nu \bar{\epsilon} \tag{I.2}$$

as if the energy were present in the form of particles (photons), with energy $h\nu$. Adding (I.1) and (I.2), one obtains equation (4).

APPENDIX II.

Classical Energy Exchange between Wave and an Electron Beam.

In a rectangular wave guide with cross-section $a \times b$ (Fig. 5), the components of the electric field are, in the TE₀₁ mode,

$$E_x = E_0 \cos\left(\frac{\pi y}{b}\right) \cos 2\pi\nu\left(t - \frac{z}{V}\right), \quad E_y = E_z = 0,$$

where V is the phase velocity

$$V = c \left[1 - \left(\frac{c}{2b\nu} \right)^2 \right]^{1/2}.$$

Thus the mean electric energy is, per unit length of the wave guide,

$$\frac{1}{8\pi} \frac{1}{4} E_0^2 ab,$$

and the mean total energy, electric and magnetic, is twice as much. This energy moves with the group velocity $U = c^2/V$. The mean number $\bar{N}$ of photons in the information cell is obtained by equating the energy flux expressed by $\bar{N}$ and expressed by E_0 . As the cell occupies a time $\Delta t = 1/\Delta\nu$, the flux per second, i.e., the power, is

$$h\nu \bar{N} \Delta\nu = \frac{1}{8\pi} \frac{1}{2} E_0^2 ab U = \frac{1}{16\pi} E_0^2 abc \left[1 - \left(\frac{c}{2b\nu} \right)^2 \right]^{1/2}. \quad (\text{II.1})$$

Consider an electron which traverses the guide on the x -axis. Assume that its velocity v is large enough to neglect its change during the transit. Thus, the number of quanta exchanged—gained or lost—will be, in the mean,

$$\bar{n} = \frac{e}{h\nu} \left| \int_{-1/2a}^{1/2a} E dx \right| = \frac{ev}{h\nu} \left| \int_{t_0}^{t_0+a/v} E dt \right|,$$

where the mean is to be taken with respect to the entrance time t_0 . A simple calculation gives

$$\bar{n} = \frac{2 \sin \theta}{\pi} \frac{ea}{\theta} \frac{e}{h\nu} E_0 \quad (\theta < \pi), \quad (\text{II.2})$$

where θ is half the “transit angle,” i.e.,

$$\theta = \pi\nu a/v.$$

The condition $\theta < \pi$ for the validity of (II.2) is always satisfied in an efficient arrangement, in fact, one must take $\theta < \frac{1}{2}\pi$.

Eliminating the field amplitude E_0 between (II.1) and (II.2), one obtains a direct relation between the photon number $\bar{N}$ and the exchange parameter $\bar{n}$:

$$\begin{aligned} \bar{n}^2 &= \frac{32 \sin^2 \theta}{\pi^2} \frac{a}{\theta^2} \frac{1}{b} \left[1 - \left(\frac{c}{2b\nu} \right)^2 \right]^{-1/2} \frac{\Delta\nu}{\nu} \bar{N} \\ &= \frac{1}{137} \frac{32 \sin^2 \theta}{\pi^3} \frac{v}{\theta} \frac{1}{\nu b} \left[1 - \left(\frac{c}{2b\nu} \right)^2 \right]^{-1/2} \frac{\Delta\nu}{\nu} \bar{N}. \end{aligned} \quad (\text{II.3})$$

As explained in the text, the coefficient of $\bar{N}$ must be made as large as possible. The depth a appears only in the form $\sin^2 \theta / \theta$, whose maximum value 0.723 is reached at $\theta = 67^\circ$, i.e. at a transit angle of 134° . The width b , on the other hand, appears in the factor

$$\frac{1}{b} \left[1 - \left(\frac{c}{2b\nu} \right)^2 \right]^{1/2} = \frac{c}{bU}.$$

The smallest admissible value of U is obtained if the group velocity is zero at the lower end of the frequency band, i.e., if frequencies below $\nu = \frac{1}{2}\Delta\nu$ are cut off. In this case

$$b = \frac{1}{2} \frac{c}{\left(\nu - \frac{1}{2}\Delta\nu \right)},$$

hence the best value of the factor in question is the reciprocal of

$$b \left[1 - \left(\frac{c}{2b\nu} \right)^2 \right]^{1/2} = \frac{c}{2\left(\nu - \frac{1}{2}\Delta\nu \right)} \left[\frac{\Delta\nu}{\nu} \left(1 - \frac{1}{4} \frac{\Delta\nu}{\nu} \right) \right]^{1/2} \approx \frac{c}{2\nu} \left(\frac{\Delta\nu}{\nu} \right)^{1/2},$$

assuming $\Delta\nu/\nu \ll 1$, i.e., the band to be narrow. Substituting these values, one obtains

$$\bar{n}^2 = \frac{1.50}{137} \frac{v}{c} \left(\frac{\Delta\nu}{\nu} \right)^{1/2} N, \quad (\text{II.4})$$

which is equation (15) of the text.

If repeated passages are used, it is evidently permissible to consider them of equal value, so long as their number P is small. It may be asked when this assumption will lead to an appreciable error. An estimate can be made by introducing the concept of the “instantaneous frequency”^{*} which will vary slowly between the limits $\nu \pm \frac{1}{2}\Delta\nu$. A somewhat long calculation, which may be omitted, gives the result that for large P ’s the factor of $\bar{n}$ is not exactly P but approximately

$$\left(\frac{\overline{\sin P\pi\delta\nu/\nu}}{\pi\delta\nu/\nu} \right) = P \left[1 - \frac{1}{6} P^2 \pi^2 \left(\frac{\overline{\delta\nu^2}}{\nu} \right) + \dots \right],$$

where $\delta\nu$ is the deviation of the instantaneous from the mean frequency. Assuming that the instantaneous frequency is uniformly distributed in the available band $\Delta\nu$, one has $\overline{\delta\nu^2} = 1/12(\Delta\nu)^2$. In the text, we have assumed that the maximum number of admissible passages is $P_{max} = \nu/\Delta\nu$. Even this gives an error of less than 14 percent, which justifies our approximate treatment.

^{*} This concept is due to Helmholtz and has been first applied in communication theory by Balh. v.d. Pol. (1930). cf. also J. R. Carson and T. C. Fry.[20]

APPENDIX III.

Energy Exchange of Electrons and Weak Fields According to Quantum Mechanics.

We apply to the problem the standard perturbation method of wave mechanics. As the beam width can be considered as small in comparison to the spatial periods of the field, it is sufficient to start from Schrödinger's one-dimensional equation with vector potential A_x ,

$$\frac{h}{2\pi i} \frac{\delta \Psi}{\delta t} = \frac{h^2}{8\pi^2 m} \frac{\delta^2 \Psi}{\delta x^2} - \frac{ieh}{2\pi mc} A_x \frac{\delta \Psi}{\delta x}. \quad (\text{III.1})$$

We put

$$A_x = A_0 \cos \omega_0 t, \quad A_y = A_z = 0, \quad \omega_0 = 2\pi\nu_0$$

which corresponds to an electric field

$$E_x = E_0 \sin \omega_0 t, \quad E_y = E_z = 0, \quad E_0 = A_0/\omega_0 c,$$

ν_0 is here the frequency of the field, which previously we have called ν , but in this calculation we will reserve ν for the frequency of the oscillators by which we will later replace the electron.

Assume that the solutions for zero field are written in the form

$$\Psi_n = \psi_n \exp(-i2\pi\epsilon_n t/h),$$

where ϵ_n is the energy of the unperturbed electron in some state “ n .” The general solution of equation (III.1) can be written in the form

$$\Psi = \sum_m c_m(t) \Psi_m.$$

With this substitution, equation (III.1) becomes

$$\sum \frac{dc_m}{dt} \Psi_m = \frac{e}{mc} \sum c_m A_x \frac{\delta \Psi_m}{\delta x}. \quad (\text{III.2})$$

Let us assume that the functions $\psi_m(x)$ are orthonormal

$$\int \psi_n^* \psi_m dx = \delta_{nm}.$$

Multiplying equation (III.2) by Ψ_r^* , integrating over the whole domain and making use of the orthonormality conditions, we obtain ordinary differential equations for the coefficients $c_r(t)$ in the form

$$\frac{dc_r}{dt} = i \frac{\pi e A_0}{mch} \sum_m c_m p_{mr} \left(\exp[i(\omega_0 + \omega_{rm})t] + \exp[-i(\omega_0 - \omega_{rm})t] \right), \quad (\text{III.3})$$

where

$$p_{mr} = \frac{h}{2\pi i} \int \psi_r^* \frac{\delta}{\delta x} \psi_m dx$$

is the matrix element of the momentum p of the electron, and

$$\omega_{rm} = \frac{2\pi}{h} (\epsilon_r - \epsilon_m)$$

is the circular frequency which corresponds to the transition $r \rightarrow m$.

Assume now that at the instant t_0 the electron is in the state $m = 0$. Thus at the start the coefficient c_0 is unity, all the others are zero. Let us also assume that during the time of interaction, i.e., during the passage of the electron, the other states c_m increase slowly enough to neglect transitions other than $0 \rightarrow r$. Thus we obtain, integrating equation (III.3) for the coefficient c_r at the time t ,

$$c_r(t) = \frac{neA_0}{mch} p_{0r} \left\{ \frac{\exp[i(\omega_0 + \omega_r)t] - \exp[i(\omega_0 + \omega_r)t_0]}{(\omega_0 + \omega_r)} - \frac{\exp[-i(\omega_0 - \omega_r)t] - \exp[-i(\omega_0 - \omega_r)t_0]}{(\omega_0 - \omega_r)} \right\}.$$

The absolute square of c_r is the probability that after the time $t - t_0$ the electron will be in the state “ r ,” having absorbed an energy corresponding to the frequency $\omega_{r0} = \omega_r$. Substituting E_0 instead of A_0 , we find finally for the probability distribution of the electron at time t the expression

$$c_r c_r^* = \left(\frac{eE_0}{mh\nu_0} \right)^2 |p_{0r}|^2 \left\{ \frac{\sin^2 \frac{1}{2}(\omega_0 + \omega_r)(t - t_0)}{(\omega_0 + \omega_r)^2} + \frac{\sin^2 \frac{1}{2}(\omega_0 - \omega_r)(t - t_0)}{(\omega_0 - \omega_r)^2} + 2 \cos 2\omega_0 t \frac{\sin^2 \frac{1}{2}\omega_r(t - t_0)}{\omega_0^2 - \omega_r^2} \right\}. \quad (\text{III.4})$$

There are several interesting features in which this differs from the classical result. Note first, that the expression is even in ω_r , hence an electron has the same probability to gain or to lose a certain amount of energy. Note also that only the last term depends on the phase of the electric field at the instant t_0 , the others depend only on the interaction time $t - t_0$. But even this last term has a frequency *twice* that of the electric field. It is rather doubtful whether this corresponds to an experimentally observable effect. Our calculation relates to weak fields, where $\bar{n}$ will be smaller than unity; but in order to observe even $\bar{n} = 1$, the electron beam must be monochromatized to such an extent that its entrance phase becomes indefinite within a whole cycle. This is of no importance for our subsequent calculations, as we will consider only the mean value of (III.4), averaged over the entrance times t_0 , in which the last term vanishes.

There remains now the problem to apply equation (III.4), which is a well-known result of wave-mechanics, to our problem of an electron traversing the oscillating field between two conducting plates. This means calculating the matrix of the (mechanical) momentum p . We

will evaluate this by a classical method, making use of the correspondence principle. This is justified in our case, as the de Broglie wavelength of the electrons is always very small with respect to the distance a between the plates.

In order to simplify the calculations, and to avoid singularities, we replace the electron by a plane density wave of total charge e (per unit cross-section)

$$\rho(x, t) = \frac{e}{\sqrt{(\pi)} v \tau} \exp[-(t - t_0 - x/v)^2/\tau] D(x). \quad (\text{III.5})$$

$D(x)$ is an “annihilation factor” which is unity inside the guide, i.e., for $-\frac{1}{2}a < x < \frac{1}{2}a$, and zero outside. Apart from this, i.e., if the electron is at more than about $v\tau$ from the walls, the distribution is Gaussian. τ is a small time, with which we will go ultimately to the limit zero. It will make the calculations easier if for the start we ignore the factor $D(x)$ and make use of it only towards the end.

We make use of the Fourier formula

$$\exp -[(t - x/v)/\tau]^2 = \sqrt{(\pi)} \tau \int_{-\infty}^{\infty} \exp[-(\pi\tau\nu)^2 - 2\pi i\nu(t - x/v)] d\nu.$$

This is to say that we decompose the electron into traveling waves of density

$$\rho_\nu d\nu = e/v \exp -(\pi\tau\nu)^2 \exp[-2\pi i\nu(t - x/v)] d\nu.$$

We have put the entrance time $t_0 = 0$, as it plays no part in the mean values which we want to calculate. In the following formulæ we will drop also the factor $\exp -(\pi\tau\nu)^2$, which cuts off the spectrum at very high frequencies, but which is practically unity in the frequency region to be investigated.

The electric field produced by the electron follows from Poisson's equation

$$\frac{\partial E}{\partial x} = 4\pi\rho.$$

Let us decompose this too into harmonic components by putting

$$E = \int_{-\infty}^{\infty} E_\nu d\nu.$$

One obtains

$$E_\nu = i \frac{2e}{\nu} \exp \left[-2\pi i\nu \left(t - \frac{x}{v} \right) \right]$$

for the progressive field waves which accompany the *free* electron. But in a plane condenser with boundaries at $x = \pm \frac{1}{2}a$, this field induces surface charges

$$\sigma_\nu \left(\pm \frac{1}{2}a \right) = \mp \frac{1}{4\pi} E_\nu \left(\pm \frac{1}{2}a \right) = \mp \frac{ie}{2\pi\nu} \exp \left[-2\pi i\nu \left(t \mp \frac{1}{2} \frac{a}{v} \right) \right],$$

which cut off the field beyond $x = \pm \frac{1}{2}a$.

We can now calculate the electric moment of the charges in the condenser, i.e., of the electron charge and the surface charges, whose sum total is zero. The Fourier component of the moment in the frequency interval ν , $d\nu$ is

$$d\nu \left\{ \int_{-\frac{1}{2}a}^{\frac{1}{2}a} x \rho_\nu dx + \frac{1}{2}a \left[\sigma_\nu \left(\frac{1}{2}a \right) - \sigma_\nu \left(-\frac{1}{2}a \right) \right] \right\} = i \frac{ev}{2\pi^2\nu^2} \sin(\pi\nu a/v) \exp(-2\pi i\nu t) d\nu.$$

We now equate this to the “standard oscillator” of electromagnetic theory, i.e., to an electron of charge e which oscillates with a frequency ν around an equal and opposite charge, with an instantaneous amplitude X_ν . This gives

$$X_\nu d\nu = i \frac{v}{2\pi^2\nu^2} \sin(\pi\nu a/v) \exp(-2\pi i\nu t) d\nu.$$

The mechanical momentum of this equivalent oscillator is $mX_\nu d\nu$. Thus the momentum is, for the frequency interval ν , $d\nu$,

$$P_\nu d\nu = \frac{mv}{\pi\nu} \sin(\pi\nu a/v) \exp(-2\pi i\nu t) d\nu. \quad (\text{III.6})$$

We cannot introduce this directly into equation (III.4), because that equation relates to discrete, steady states. In that case one has only to take the absolute square of the matrix elements to form the quantities $|p_{or}|^2$ which figure in equation (III.4). But here we have to do with continuous states, and we must also remember that the oscillators, according to (III.6), replace the electron during the finite transit time only. In order to obtain the equivalents of the “oscillator strengths” $|p_{or}|^2$, we must now multiply the momentum amplitudes $P_\nu d\nu$ with conjugate amplitudes $P_\mu^* d\mu$ which belong to another interval μ , $d\mu$; we must integrate over the whole domain of the μ , and finally average over the transit time a/v . It is to say that we must calculate, instead of $|p_{or}|^2$,

$$|p_{or}|^2 d\nu = 2 \frac{v}{a} d\nu \int_{-\frac{1}{2}a/v}^{\frac{1}{2}a/v} dt \int_{-\infty}^{\infty} P_\nu P_\mu^* d\mu. \quad (\text{III.7})$$

We had to add the factor two, because in wave mechanics ν is considered as a positive quantity, while in classical theory it runs from $-\infty$ to $+\infty$. Substituting from (III.6),

$$|p_{or}|^2 = 2 \left(\frac{mv}{\pi} \right)^2 \frac{v}{a} \frac{\sin \pi\nu a/v}{\nu} \int_{-\frac{1}{2}a/v}^{\frac{1}{2}a/v} dt \int_{-\infty}^{\infty} \frac{\sin \pi\mu a/v}{\mu} \exp[2\pi i(\mu - \nu)t] d\mu.$$

After integration with respect to t , the double integral becomes

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin \pi\mu a/v}{\mu} \sin \frac{\pi(\mu - \nu)a/v}{\mu - \nu} d\mu.$$

With the substitutions $x = (2\mu/v) - 1$ and $\theta = \pi\nu a/v$, this is transformed into

$$\frac{2}{\pi\nu} \int_{-\infty}^{\infty} \frac{\sin \frac{1}{2}\theta(1+x) \sin \frac{1}{2}\theta(1-x)}{1-x^2} dx = \frac{2}{\pi\nu} \int_0^\infty \frac{\cos \theta x - \cos \theta}{1-x^2} dx = \frac{1}{\nu} \sin \theta,$$

thus finally

$$|p_{or}|^2 d\nu = 2 \left(\frac{m\nu}{\pi} \right)^2 \frac{v \sin^2 \theta}{a \nu^2} d\nu. \quad (\text{III.8})$$

This means simply that during the transit time a/v the coherence region of any frequency ν is $\nu \pm v/2a$, with a bandwidth v/a , which is the reciprocal of the transit time. A frequency ν cooperates with any frequency μ inside this band in producing the oscillator strength.

Substituting (III.8) instead of $|p_{or}|^2$ into equation (III.4), where we drop the last term whose mean value is zero, and replacing $t - t_0$ by the transit time a/v , we obtain

$$|c_\nu|^2 d\nu = \frac{1}{h} \left(\frac{2\pi e^2}{hc} \right) \frac{E_0^2 acv}{4\pi^3 \nu_0^2} \frac{d\nu}{\nu^2} \sin^2(\pi \nu a/v) \left\{ \frac{\sin^2 \pi(\nu_0 + \nu)a/v}{[(\nu_0 + \nu)a/v]^2} + \frac{\sin^2 \pi(\nu_0 - \nu)a/v}{[(\nu_0 - \nu)a/v]^2} \right\} \quad (\text{III.9})$$

This is the probability for an electron to be, after the passage, in a state ν , $d\nu$, that is to say, to have absorbed the energy $h\nu$. Thus the energy spectrum of the originally monokinetic electrons after the passage through the field is of the form

$$S(\theta) = \frac{\sin^2 \theta}{\theta^2} \left\{ \frac{\sin^2(\theta + \theta_0)}{(\theta + \theta_0)^2} + \frac{\sin^2(\theta - \theta_0)}{(\theta - \theta_0)^2} \right\}, \quad (\text{III.10})$$

where the frequencies (and energies) are expressed by the half transit angles

$$\theta = \pi \nu a/v, \quad \theta_0 = \pi \nu_0 a/v.$$

This function is shown in Fig. 8 for a few values of θ_0 . Experimental *checking*, though difficult, may not be impossible (cf. Macdonald and Kompfner, 1949[21]).

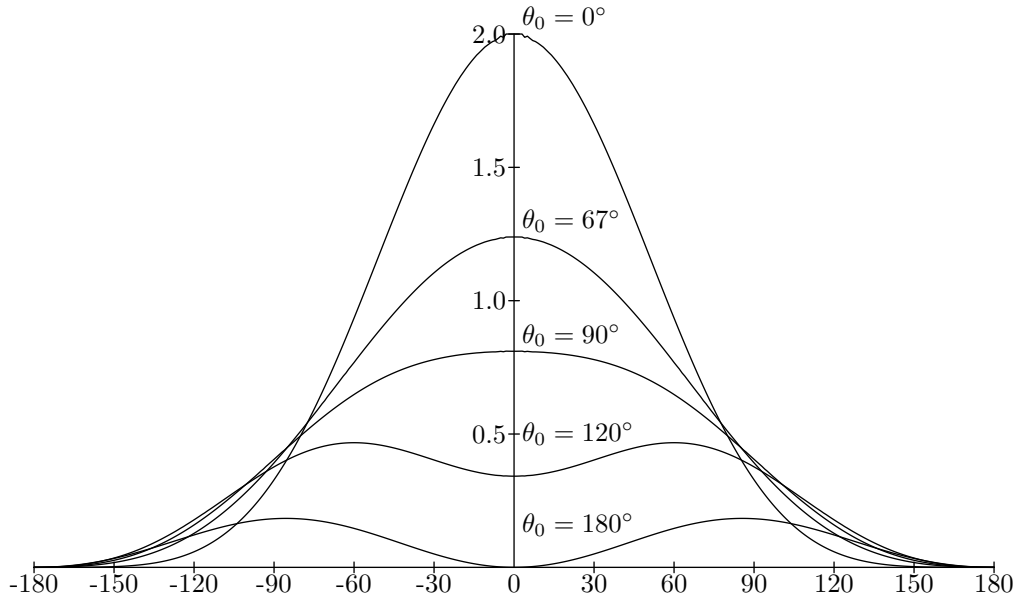
The mean energy exchanged with the field, lost or gained, is by (III.10)

$$\bar{n} h \nu_0 = \int_0^\infty h \nu |c_\nu|^2 d\nu = \frac{1}{137} \frac{E_0^2 acv}{4\pi^2 \nu_0^2} F(\theta_0), \quad (\text{III.11})$$

where $F(\theta_0)$ is the integral

$$F(\theta_0) = \int_0^\infty \frac{d\theta}{\theta} \left[\frac{\sin^2(\theta + \theta_0)}{(\theta + \theta_0)^2} + \frac{\sin^2(\theta - \theta_0)}{(\theta - \theta_0)^2} \right].$$

Fig. 8



The function $S(\theta)$, equation (III.10), which gives the energy distribution of electrons after the transit. Gains and losses have equal probabilities. The “resonance point” at which the electron has gained or lost one quantum $h\nu_0$ of the oscillating field is at $\theta = \theta_0$.

For $\theta = 67^\circ$, which was the most favorable case in the classical theory, the value of the integral is found to be 1.22. Substituting this into equation (III.11) and expressing, as before the field E_0 by the number of photons, $\bar{N}$, by means of equation (II.1), and finally giving the wave guide the same width b which was found the best in the classical case, one finds

$$\bar{n} = \frac{1.0}{137} \frac{v}{c} \left(\frac{\Delta\nu}{\nu} \right)^{1/2} \bar{N},$$

where we have again written ν for the field frequency, instead of ν_0 . This is the energy exchange in weak fields, discussed in the text.

APPENDIX IV.

Conditions for Optimum Exploration of Electric Fields by Electrons.

As explained in the text, we adopt as the criterion of optimum measurement the condition that

$$\frac{\overline{(nM - CN^{1/2})^2}}{\overline{(nM)^2}},$$

must be as small as possible. The proportionality factor C is given by the condition

$$\overline{(nM - CN^{1/2})} = 0,$$

or, as we are dealing here only with the “classical” case, in which the fluctuations are relatively small, $C = \overline{nM} (\overline{N})^{-1/2}$. As the fluctuations of n , the energy exchange parameter, and of M , the electron number, can be supposed as independent of one another, we can write $\overline{nM} = \overline{n} \overline{M}$.

Let us now first calculate

$$\begin{aligned} \overline{n^2 M^2} &= \overline{[(\overline{n} + \delta n)(\overline{M} + \delta M)]^2} = \overline{(\overline{n} \overline{M} + \overline{n} \delta M + \overline{M} \delta n)^2} \\ &= \overline{n^2} \overline{M^2} + \overline{n^2} \overline{\delta M^2} + \overline{M^2} \overline{\delta n^2} = \overline{n^2} \overline{M^2} + \overline{n^2} \overline{M} + \overline{M^2} \overline{n}. \end{aligned}$$

Here we have neglected the fluctuation term of the highest order, and we have assumed

$$\overline{\delta n \delta M} = 0, \quad \overline{\delta n^2} = \overline{n}, \quad \overline{\delta M^2} = \overline{M}.$$

The first is the assumption of the relative independence of the fluctuations of n and M . The second is the assumption that the absorption of a quantum is independent of those previously absorbed or lost, which is certainly admissible so long as the electron velocity is not changed appreciably. The third assumption means shot effect without space charge smoothing. This is a logical assumption in a theory in which each electron is considered to interact singly with the field. With these formulæ, we can now calculate

$$\begin{aligned} \overline{(nM - CN^{1/2})^2} &= \overline{n^2 M^2} - 2C \overline{(\overline{n} + \delta n)(\overline{M} + \delta M)} \overline{(N^{1/2} + \frac{1}{2} \delta N (N^{-1/2}))} \\ &\quad + C^2 \overline{(N^{1/2} + \frac{1}{2} \delta N (N^{-1/2}))^2} \\ &= (\overline{n} \overline{M} - C \overline{N^{1/2}})^2 + \overline{n^2} \overline{M} + \overline{M^2} \overline{n} + \frac{1}{4} C^2 \frac{\overline{\delta N^2}}{\overline{N}}, \end{aligned} \tag{IV.1}$$

where the first term is zero.

In order to calculate the last term, we consider the fluctuation δN of the photons as the sum of two independent components δN_1 and δN_2

$$\delta N = \delta N_1 + \delta N_2,$$

of which the first represents the “natural” fluctuations, subject to the law $\overline{\delta N_1^2} = \overline{N}$, and the second is due to the exploring electrons, i.e., it is the excess of emitted over absorbed photons. Let M_1 be the number of beam electrons which pass through the field in a retarding phase, M_2 the number of accelerated electrons. As in the mean $\overline{M_1} = \overline{M_2} = \frac{1}{2}\overline{M}$, and as in the mean each electron produces or absorbs $\overline{n}$ photons, we can write

$$\delta_2 N + \overline{n}(M_1 - M_2) = \overline{n}(\delta M_1 - \delta M_2),$$

and

$$\overline{\delta_2 N^2} = \overline{n^2}(\overline{\delta M_1^2} + \overline{\delta M_2^2}) = \overline{n^2}(\overline{M_1} + \overline{M_2}) = \overline{n^2}\overline{M},$$

where we have assumed $\overline{\delta M_1 \delta M_2} = 0$, in accordance with our assumption of a normal, entirely random shot effect.

We now substitute these results into equation (IV.1), and replacing C^2 by its value $\overline{n^2}\overline{M}^2\overline{N}$, we obtain for the mean square relative error in the measurement of the amplitudes

$$\frac{\overline{(nM - CN^{1/2})^2}}{\overline{nM^2}} = \frac{1}{4\overline{N}} + \frac{1}{\overline{n}} + \overline{n}\left(\frac{1}{\overline{nM}} + \frac{\overline{nM}}{4\overline{N^2}}\right). \quad (\text{IV.2})$$

This is equation (24), discussed in the text.

It is not so easy to give an exact analysis of the errors in the measurement of phase, because devices for the running measurement of phase, with uniform error, are by no means simple. But the order of magnitude can be seen immediately if we apply the uncertainty principle to the exploring electrons, in the form

$$\delta_\epsilon \delta t \approx 1.$$

δt is the uncertainty in the time at which the electron passes some fixed plane, e.g., a florescent screen on which the waveform is recorded. The uncertainty in the energy ϵ is, assuming that the electrons were monokinematic at the start,

$$\delta\epsilon = h\nu\delta n,$$

n being the number of quanta which it may have absorbed from or lost to the field. But as δn is of the order $\overline{n}^{1/2}$, we obtain

$$\nu\delta t \approx (\overline{n})^{-1/2}.$$

In words, the oscillogram of the signal is traced with a spot whose mean width in the direction of time is a fraction $1/\sqrt{\overline{n}}$ of a cycle. Thus one can determine the phase with the same order of relative error as the amplitude.